Entropy Rate of Diffusion Processes on Complex Networks

Jesús Gómez-Gardeñes[1,2] and Vito Latora[3]

[1]*Scuola Superiore di Catania, Via S. Paolo 73, 95123 Catania, Italy*
[2]*Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, Zaragoza 50009, Spain*
[3]*Dipartimento di Fisica e Astronomia, Università di Catania and INFN, Via S. Sofia 64, 95123 Catania, Italy*

(Dated: February 2, 2008)

The concept of entropy rate for a dynamical process on a graph is introduced. We study diffusion processes where the node degrees are used as a local information by the random walkers. We describe analitically and numerically how the degree heterogeneity and correlations affect the diffusion entropy rate. In addition, the entropy rate is used to characterize complex networks from the real world. Our results point out how to design optimal diffusion processes that maximize the entropy for a given network structure, providing a new theoretical tool with applications to social, technological and communication networks.

PACS numbers: 02.50.Ga,89.75.Fb,89.75.Hc

Entropy is a key concept in statistical thermodynamics [1], in the theory of dynamical systems [2], and in information theory [3]. In the realm of complex networks [4, 5], the entropy has been used as a measure to characterize properties of the topology, such as the *degree distribution* of a graph [6], or the *shortest paths* between couples of nodes (with the main interest in quantifying the information associated with locating specific addresses [7], or to send signals in the network [8]. Alternatively, various authors have studied the entropy associated with *ensembles of graphs*, and provided, via the application of the maximum entropy principle, the best prediction of network properties subject to the constraints imposed by a given set of observations [9, 10, 11]. An approach of this type plays the same role in networks as is played by the Boltzmann distribution in statistical thermodynamics [1].

The main theoretical and empirical interest in the study of complex networks is in understanding the relations between structure and function. Besides, many of the interaction dynamics that takes place in social, biological and technological systems can be analyzed in terms of *diffusion processes* on top of complex networks, *e.g.* data search and routing, information and disease spreading [4, 5]. It is therefore of outmost importance to relate the properties of a diffusion process with the structure of the underlying network.

In this Letter, we show how to associate an *entropy rate* to a diffusion process on a graph. In particular, we consider processes such as biased random walks on the graph that can be represented as ergodic Markov chains. In this context, the entropy rate is a quantity more similar to the Kolmogorov-ÂSinai entropy rate of a dynamical system [12, 13], than to the entropy of a statistical ensemble [1, 4]. Differently from the network entropies previously defined, the entropy rate of diffusion processes depends both on the dynamical process (the kind of bias in the random walker) and on the graph topology. We provide the analytical expression that describes the entropy rate in scale-free networks as a function of the bias in the walk, and of the degree distribution and correlations. We show how the values of the entropy rate can provide useful information to characterize diffusion processes in real-world networks. In particular, a maximum value of entropy is found for different types of the bias in the diffusion processes, depending on the network structure.

Let us consider a connected undirected graph with N nodes (labelled as $1, 2, ..., N$) and K links, described by the adjacency matrix $A = \{a_{ij}\}$ We limit our discussion to diffusion processes on the graph that can be represented as *Markov chains* [3]. In particular, we consider the case of biased random walks in which, at each time step, the walker at node i chooses one of the first neighbors of i, let say j, with a probability proportional to the power α ($\alpha \in \mathbb{R}$) of the degree k_j. Such biased random walk corresponds to a time-invariant (the rule does not change in time) Markov chain with a *transition probability matrix* Π, with elements:

$$\pi_{ji} = \frac{a_{ij} k_j^{\alpha}}{\sum_j a_{ij} k_j^{\alpha}} \qquad (1)$$

Notice that Π depends on either the graph topology and the kind of stochastic process we are considering. The exponent α allows to tune the dependence of the diffusion process on the nodes' degree. When $\alpha \neq 0$ we are introducing in the random movement of the particle a bias towards high- ($\alpha > 0$) or low-degree (when $\alpha < 0$) neighbors. On the other hand, when $\alpha = 0$ the standard (unbiased) random walk is recovered. Since the walker must move from a node to somewhere, we have $\sum_j \pi_{ji} = 1$, thus Π is a stochastic matrix. If $w_i(t)$ is the probability that the random walker is at node i at time t (with $\sum_{i=1}^{N} w_i(t) = 1 \; \forall t$), then the probability $w_j(t+1)$ of its being at j one step later is: $w_j(t+1) = \sum_i \pi_{ji} w_i(t)$. Writing the probabilities $w_i(t)$ as a N-dimensional column vector $\mathbf{w}(t) = (\, w_1(t), w_2(t) \ldots w_N(t)\,)^{\top}$, the rule of the walk can be expressed in matricial form as: $\mathbf{w}(t+1) = \Pi \mathbf{w}(t)$. In the case of an undirected and connected network, the Perron-Frobenius theorem [14] assures that the dynamics described by Eq. (1) is an *ergodic Markov chain* [3]. This means that the Markov chain has a unique *stationary distribution* $\mathbf{w}^*$, such that $\lim_{t\to\infty} \Pi^t \mathbf{w}(0) = \mathbf{w}^*$ for any initial distribution $\mathbf{w}(0)$. In other words, any initial distribution of the random walker over the nodes of the graph will converge, under the dynamics of Eq. (1), to the same distribution $\mathbf{w}^*$.

The dynamical properties of the above diffusion processes

over the graph can be accounted by evaluating the *entropy rate* of the associated Markov chain that, in the case of an ergodic Markov chain, is given by [3]:

$$h = -\sum_{i,j} \pi_{ji} \cdot w_i^* \ln(\pi_{ji}) \tag{2}$$

The value of h measures how the entropy of the biased random walk grows with the number of hops. This means that we can practically represent the typical sequences of length n generated by the diffusion process by using approximately $n \cdot h$ information units. In different words, h measures the spreading of a set of independent random walkers, in terms of number of visited nodes.

To evaluate h for a given graph we need to calculate the stationary probability distribution $\mathbf{w}^*$. For this purpose, we consider the probability $W_{i\to j}(t)$ of going from node i to node j in t time steps,

$$W_{i\to j}(t) = \sum_{j_1,j_2,\ldots,j_{t-1}} \pi_{i,j_1} \cdot \pi_{j_1,j_2} \cdot \ldots \cdot \pi_{j_{t-1},j} \; . \tag{3}$$

Since the network is undirected we have $a_{ij} = a_{ji}\ \forall i,j$. Hence, the relation between the two probabilities $W_{i\to j}(t)$ and $W_{j\to i}(t)$ can be written as:

$$c_i k_i^\alpha W_{i\to j}(t) = c_j k_j^\alpha W_{j\to i}(t) \; , \tag{4}$$

where $c_i = \sum_j a_{ij} k_j^\alpha$. The above relation implies that for the stationary distribution $\mathbf{w}^*$ the equation $c_i k_i^\alpha w_j^* = c_j k_j^\alpha w_i^*$ holds, and hence $\mathbf{w}^*$ reads:

$$w_i^* = \frac{c_i k_i^\alpha}{\sum_l c_l k_l^\alpha} \; . \tag{5}$$

By plugging expressions (1) and (5) into the definition of entropy (2), we finally get a closed form for the entropy rate of degree-biased random walks on the graph:

$$h = \frac{\sum_i k_i^\alpha \sum_j a_{ij} k_j^\alpha \ln(k_j^\alpha) - \sum_i k_i^\alpha c_i \ln(c_i)}{\sum_i c_i k_i^\alpha} \; . \tag{6}$$

We notice that h depends on the the kind of bias in the random walker and also on the graph topology. In the following, we first evaluate analytically the entropy rate of unbiased and biased random walks on scale-free (SF) graphs with a power-law degree distribution $P_k \sim k^{-\gamma}$, and $\gamma > 2$ [4, 5]. Then, we study the entropy in networks from the real world.

Unbiased Random Walks.- In the particular case $\alpha = 0$, the transition probability reads $\pi_{ji} = a_{ij}/k_i$, and the stationary distribution is easily obtained as: $w_i^* = \frac{k_i}{2K}$. Substituting this expression in Eq. (2) and changing the sum over node indexes into a sum over degree classes, we can write the entropy rate of a unbiased random walk on a network with degree distribution P_k as:

$$h = \frac{N}{2K} \sum_k k P_k \ln(k) = \frac{\langle k \ln(k) \rangle}{\langle k \rangle} \; . \tag{7}$$

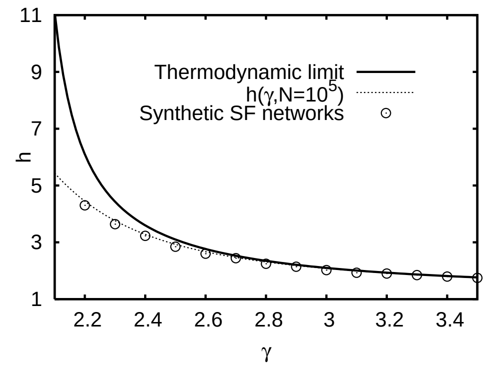

FIG. 1: Entropy rate h of unbiased random walks on SF networks with $N = 10^5$ nodes as a function of the exponent γ of the degree distribution. Numerical results (circles) are compared with the two analytical curves corresponding to Eq. (8) (dashed line) and to the limit $N \to \infty$ (solid line).

In the case of SF networks of size N, the value of h can be easily expressed as a function of γ and N taking into account that the maximum degree of the network is $k_{max} \sim k_0 N^{1/(\gamma-1)}$, with k_0 being the minimum degree of a node. From Eq. (7), and approximating k as a continuum variable, we get:

$$h(\gamma, N) = \ln(k_0) + \frac{1}{\gamma-2} + \frac{N^{\frac{2-\gamma}{\gamma-1}} \ln(N)}{(\gamma-1)(N^{\frac{2-\gamma}{\gamma-1}} - 1)} \; . \tag{8}$$

The above expression diverges for SF networks when $\gamma \to 2$. Conversely, when $\gamma > 2$ SF networks have a finite entropy in the thermodynamic limit: $h(\gamma) = \ln(k_0) + \frac{1}{\gamma-2}$.

In order to check the analytical results we have constructed ensembles of 10^2 SF networks with $N = 10^5$ nodes and different values of γ. We have obtained numerically the stationary distribution $\mathbf{w}^*$, and computed the entropy directly from Eq. (2). The results, averaged over the ensemble of networks, are reported in Fig. 1 as a function of γ. We notice a good agreement between numerics and Eq. (8).

Biased Random Walks.- Let us now concentrate on degree-biased diffusion ($\alpha \neq 0$). In this case, the entropy rate of Eq. (6) can be re-written by changing again the sums over node indexes into sums over degree classes, as:

$$h = -\frac{\sum_k k^\alpha P_k \left(C_k \ln(C_k) - \sum_{k'} \alpha k'^{\alpha} P_{k',k} \ln(k') \right)}{\sum_k C_k k^\alpha P_k} \tag{9}$$

where $C_k = k \sum_{k'} k'^{\alpha} P_{k',k}$, and $P_{k',k}$ is the conditional probability that a link from a node of degree k ends in a node with degree k'. We notice that the entropy rate of biased random walks depends on the degree distribution of the network, P_k, and on the conditional probabilities $P_{k',k}$. In the particular case of a network with no degree-degree correlations we can write $P_{k',k} = k P_k / \langle k \rangle$, and the expression for the entropy

reduces to:

$$h = (1-\alpha)\frac{\langle k^{\alpha+1}\ln(k)\rangle}{\langle k^{\alpha+1}\rangle} + \ln\left(\frac{\langle k^{\alpha+1}\rangle}{\langle k\rangle}\right) . \qquad (10)$$

This expression only depends on the degree distribution of the network. For SF networks, we get in the the continuum-degree approximation:

$$\begin{aligned} h(\gamma,\alpha,N) &= \frac{1-\alpha}{\gamma-\alpha-2} + \frac{(1-\alpha)N^{\frac{\alpha+2-\gamma}{\gamma-1}}\ln(N)}{(\gamma-1)(N^{\frac{\alpha+2-\gamma}{\gamma-1}}-1)} \\ &+ \ln\left[\frac{k_0(\gamma-2)(N^{\frac{\alpha+2-\gamma}{\gamma-1}}-1)}{(\gamma-\alpha-2)(N^{\frac{2-\gamma}{\gamma-1}}-1)}\right] . \qquad (11) \end{aligned}$$

When $N \to \infty$, the entropy rate in SF networks with $\gamma < 2+\alpha$ diverges as $h \sim \ln(N)$. On the other hand, when $\gamma > 2+\alpha$, the entropy rate in the limit $N \to \infty$ is finite and equal to:

$$h(\gamma,\alpha) = \frac{1-\alpha}{\gamma-\alpha-2} + \ln\left[\frac{k_0(\gamma-2)}{\gamma-\alpha-2}\right] . \qquad (12)$$

Such an expression, valid in infinite size limit, shows a monotone growth of the entropy $h(\gamma,\alpha)$ with the degree-bias α, with h tending to infinity as $\alpha \to (\gamma-2)^-$. More interestingly, the entropy rate in finite networks, Eq. (11), shows a single maximum at a value of α that depends on γ. This result is a consequence of the interplay between diffusion process and network topology. It indicates that, for a given network, is possible to maximize the entropy of the process by opportunely tuning the bias α of the walker.

To check the above analytical expressions we have computed numerically the entropy rate of degree-biased random walkers on computer-generated uncorrelated SF networks, as we did for the unbiased case. In Fig. 2.a we report the entropy rate as a function of the degree bias α for SF networks of size $N = 10^5$. In Fig. 2.b we show the scaling of h with the system size N, in SF network with $\gamma = 3$. In both cases Eq. (11) is in good agreement with the numerical results reproducing the qualitative behavior of h as a function of α (being the global maximum of h well reproduced) and N (being both the divergence of h, for $\alpha > \gamma - 2$, and the asymptotic finite value of h, when $\alpha < \gamma - 2$, correctly reproduced) [15].

Real Networks.- Up to now, we focused on the entropy rate of biased random walks on SF networks. However, real networks are not perfect scale-free and, more importantly, show additional important structural properties such as degree-degree correlations, motifs and community structures [4, 5]. Now we propose to characterize a real network by studying different diffusion processes on top of it, and finding the optimal value of the bias that maximizes the entropy. As reference system, we compare the entropy rate h of α-biased random walks on the network, with the entropy rate h^{Rand} obtained, from Eq. (10), for a randomized version of the network, with the same degree sequence of the real one [16]. For this purpose, we have analyzed 10 different networks reported in Table I, corresponding to thre different functional classes where diffusion of data, rumors, viruses and

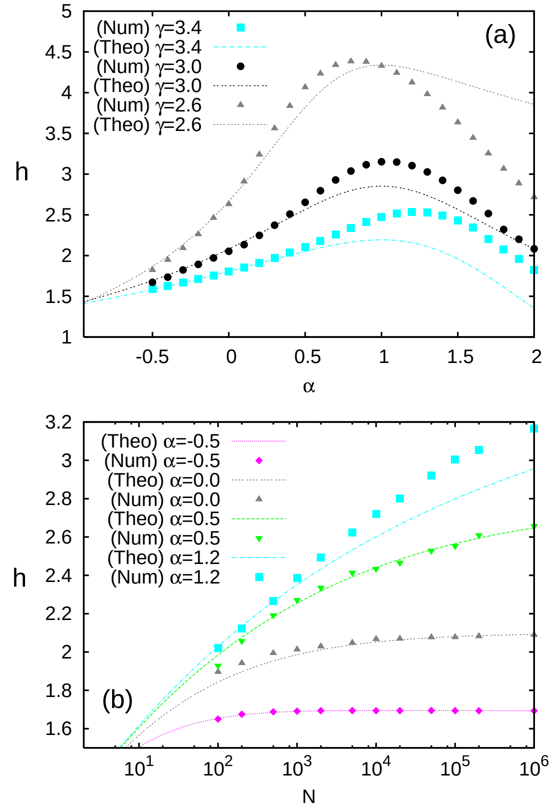

FIG. 2: (color online). (a) Entropy rate h, as a function of α, for α-biased random walks on SF networks with $N = 10^5$ nodes and $\gamma = 2.6, 3, 3.4$. Symbols represent the values of h found numerically, while the lines are the corresponding analytical predictions $h(\gamma,\alpha,N)$ of Eq. (11). **(b)** Entropy rate h for α-biased random walks on SF networks with $\gamma = 3$, as a function of the system size N and for several values of α. Again, symbols are the results of numerical simulations, while the lines correspond to Eq. (11).

diseases, takes place, namely *(i)* transportation, *(ii)* technological/communication and *(iii)* social networks.

In Fig. 3 we report, for six of the networks, the results obtained as a function of the bias parameter α. Two different behaviors emerge clearly for $\alpha > 0$ [26], namely the entropy of the real network h is either larger or smaller than h^{rand} for all the range of positive values of α. In table I we summarize this result by reporting the ratio h/h^{Rand} for $\alpha = 1$ (linear bias). We found that social networks have always $h > h^{Rand}$, while the other networks have $h < h^{Rand}$, with the exception of Internet routers. This difference in the entropy rate has its roots mainly on the different types of degree-degree correlations of the network, and points out that assortativity facilitates the spread of the diffusion. The optimal degree-bias, α^{opt} that produces the maximal entropy rate is also reported in Table I. The results indicate that for assortative networks (*e.g.* social networks) the maximal entropy rate is obtained with a super-linear diffusion, while for dissassortative networks α^{opt} is located in the sub-linear bias region.

TABLE I: Properties of the 10 real networks analyzed. N is the number of nodes in the giant connected component, $\langle k \rangle$ is the average degree. The ratio of entropy rates, h/h^{Rand}, is reported for a linear ($\alpha = 1$) degree-biased random walk. Finally we report the optimal value of α that maximizes h for each network.

network	**ref.**	N	$\langle k \rangle$	h/h^{Rand}	α^{opt}
U.S. Airports	[17]	500	11.92	0.964	0.8
Internet routers	[18]	228263	2.80	1.191	1.7
Internet A.S.	[18]	1174	4.19	0.662	0.6
WWW	[19]	325729	6.70	0.867	0.9
P2P	[20]	79939	4.13	0.613	0.7
Sci. Coll. (cond-mat)	[21]	12722	6.28	1.091	1.5
Sci. Coll. (astro-ph)	[22]	13259	18.62	1.071	1.5
U.S. patents	[23]	230686	4.81	1.113	1.5
E-mail	[24]	1133	9.62	1.019	1.2
P.G.P	[25]	10680	4.56	1.176	1.3

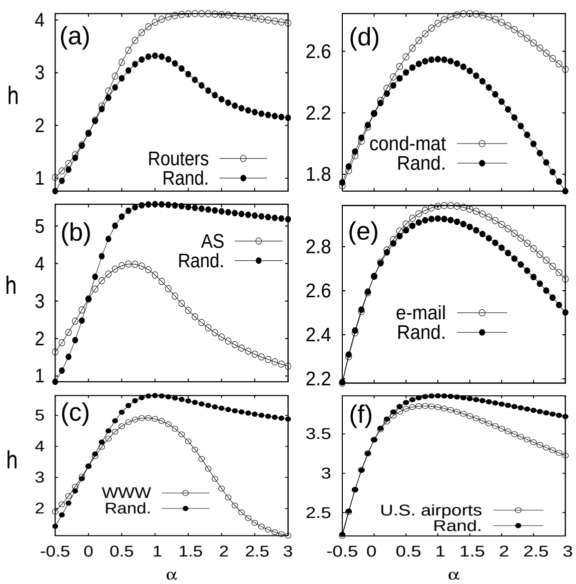

FIG. 3: Entropy rate h for α-biased random walks on six of the networks in Table I (filled circles). Such entropy rate is compared with that obtained for the randomized version of the network (full circles). Both entropy rates are shown as a function of the degree-bias parameter α.

Summing up, in this Letter, we have introduced the entropy rate of degree-biased random walks on networks, a measure that is particularly suited to capture the interplay between network structure and diffusion dynamics. We have studied the dependence of the entropy rate with the topology of synthetic and real networks, in particular with the heterogeneity of degree distributions and the nature of the degree-degree correlations. The results indicate how it is possible to tune the bias in the random walk in order to maximize the entropy rate on a given topology. The method introduced can find useful applications to cases where diffusion in complex networks is the mechanism at work, such as in the search of efficient algorithms for data search in the WWW, in the improvement of information dissemination in social networks, or in the design of large impact virus/antivirus spreading in computer networks. The approach adopted here can be easily generalized to other types of diffusion processes and to more general network topologies, such as weighted graphs and also, with some appropriate modifications to directed and unconnected graphs.

The authors are indebted to M. Chavez and H. Touchette for many helpful discussions on the subject.

[1] R. Balian, From Microphysics to Macrophysics (Springer-Verlag, New York, 1991)
[2] C. Beck, F. Schlögl, Thermodynamics of chaotic systems (Cambridge University Press, Cambridge, 1993)
[3] T. M. Cover and J. A. Thomas, Elements of Information Theory (Wiley, 1991)
[4] M.E.J. Newman, SIAM Review **45**, 167 (2003).
[5] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez , D.-U. Hwang, Phys. Rep. **424**, 175 (2006).
[6] R. Ferrer, I. Cancho and R. V. Solé, Lect. Notes in Physics **625**, 114 (2003)
[7] M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen, Phys. Rev. Lett. **94**, 028701 (2005).
[8] A. Trusina, M. Rosvall, K. Sneppen, Phys. Rev. Lett. **94** 238701 (2005).
[9] G. Bianconi, "*The Entropy of randomized network ensembles*", arXiv:0708.0153, in press in Europhys. Lett.
[10] P. Minnhagen and S. Bernhardsson, Chaos **17**, 026117 (2007).
[11] J. Park and M.E.J. Newman, Phys. Rev. **E70**, 066117 (2004).
[12] A. N. Kolmogorov, Dokl. Akad. Nauk SSSR **119**, 861 (1958); **124**, 754 (1959).
[13] V. Latora, M. Baranger, Phys. Rev. Lett. **82**, 520 (1999).
[14] F. R. Gantmacher, The Theory of Matrices, Volume 2 (Chelsea Publishing Company, New York, 1959)
[15] Note that the effect of the continuum-degree approximation, made in the calculation, become more evident as α grows, because for large α the discreteness effects in k are amplified.
[16] S. Maslov, K. Sneppen, Science **296**, 910 (2002).
[17] V. Colizza, R. Pastor-Satorras and A. Vespignani, Nature Physics **3**, 276 (2007).
[18] A. Vázquez, R. Pastor-Satorras and A. Vespignani, Lect. Notes in Physics **650**, 425 (2004).
[19] R. Albert, H. Jeong and A.L. Barabási, Nature **401**, 130 (1999).
[20] F. Wang, Y. Moreno and Y. Sun, Phys. Rev. E **73**, 036123 (2006).
[21] M.E.J. Newman, Phys. Rev. E **64** 016131 (2001); Phys. Rev. E **64** 016132 (2001).
[22] A. Cardillo, Degree Thesis: "*Topological Analysis of Scientific Coauthorship Networks*" (Univ. of Catania, Catania, 2006).
[23] B.H. Hall, A.B. Jaffe and M. Tratjenberg, NBER Working Paper 8498 (2001).
[24] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt and A. Arenas, Phys. Rev. E **68**, 065103(R) (2003).
[25] M. Boguña, R. Pastor-Satorras, A. Diaz-Guilera and A. Arenas, Phys. Rev. E **70**, 056122 (2004).
[26] Note that for $\alpha = 0$ both the real network and the randomized ensemble have the same value of h since in this limit Eq. (10) (or alternatively Eq. (7)) is exact and it only depends on the degree sequence of the network.